

VERTICAL FLOATING GATE TRANSISTOR

STATEMENT OF RELATED APPLICATIONS

[0001] This application is a divisional of U.S. Patent Application Serial No. 10/229,476 (allowed), filed August 28, 2002 and titled, "VERTICAL FLOATING GATE TRANSISTOR" which is commonly assigned and incorporated by reference in its entirety herein.

FIELD OF THE INVENTION

[0002] The present invention relates generally to non-volatile memory cells and in particular the present invention relates to flash memory cells.

BACKGROUND OF THE INVENTION

[0003] Memory devices are available in a variety of styles and sizes. Some memory devices are volatile in nature and cannot retain data without an active power supply. A typical volatile memory is a DRAM which includes memory cells formed as capacitors. A charge, or lack of charge, on the capacitors indicate a binary state of data stored in the memory cell. Dynamic memory devices require more effort to retain data than non-volatile memories, but are typically faster to read and write.

[0004] Non-volatile memory devices are also available in different configurations. For example, floating gate memory devices are non-volatile memories that use floating gate transistors to store data. The data is written to the memory cells by changing a threshold voltage of the transistor and is retained when the power is removed. The transistors can be erased to restore the threshold voltage of the transistor. The memory may be arranged in erase blocks where all of the memory cells in an erase block are erased at one time. These non-volatile memory devices are commonly referred to as flash memories.

[0005] The non-volatile memory cells are fabricated as floating gate memory cells and include a source region and a drain region that is laterally spaced apart from the source region to form an intermediate channel region. The source and drain regions are formed in a common horizontal plane of a silicon substrate. A floating gate, typically made of doped polysilicon, is disposed over the channel region and is electrically isolated from the other cell elements by oxide. For example, gate oxide can be formed between the floating gate and the channel region. A control gate is located over the floating gate and is can also made of doped polysilicon. The control gate is electrically separated from the floating gate by another dielectric layer. Thus, the floating gate is “floating” in dielectric so that it is insulated from both the channel and the control gate.

[0006] As semiconductor devices get smaller in size, designers are faced with problems associated with the production of memory cells that consume a small enough amount of surface area to meet design criteria, yet maintain sufficient performance in spite of this smaller size.

[0007] For the reasons stated above, and for other reasons stated below which will become apparent to those skilled in the art upon reading and understanding the present specification, there is a need in the art for an improved non-volatile memory cell.

SUMMARY OF THE INVENTION

[0008] The above-mentioned problems with non-volatile memory cells and other problems are addressed by the present invention and will be understood by reading and studying the following specification.

[0009] In one embodiment, a pair of floating gate transistors comprises a pillar of silicon vertically extending from a substrate, an implanted drain region located in a top region of the pillar; and a pair of source regions implanted in the substrate. The pair of source regions are located on opposite sides of the pillar. First and second floating gates are located on opposite sides of the pillar and insulated from the pillar by gate

oxide. First and second control gates are located on opposite sides of the pillar and insulated from the first and second floating gates.

[0010] In another embodiment, a non-volatile memory array comprises a pillar of silicon vertically extending from a substrate, an implanted drain region located in a top region of the pillar; and a pair of source regions implanted in the substrate. The pair of source regions are located on opposite sides of the pillar. First and second floating gates are located on opposite sides of the pillar and insulated from the pillar by gate oxide. First and second control gates are located on opposite sides of the pillar and insulated from the first and second floating gates. The first and second control gates are fabricated as parallel polysilicon rails, and the pair of source regions are fabricated as parallel source implant rails generally extending under the first and second control gates. A drain contact vertically extends from the top region of the pillar, and first and second source contacts vertically extend from the pair of source regions. A source conductor is coupled to a top of the first and second source contacts. Finally, a drain conductor is coupled to a top of the drain contact. The source and drain conductors extend in a horizontal direction generally perpendicular to the control gates.

[0011] A method of fabricating non-volatile memory cells comprises forming a pillar of silicon vertically extending above a substrate, implanting a drain region in a top of the pillar, and implanting first and second source regions in the substrate and adjacent to the pillar. The first and second source regions are located on opposite sides of the pillar. The method further comprises depositing a gate oxide on at least vertical sides of the pillar facing the first and second source regions, forming floating gates adjacent to and on opposite sides of the pillar, and forming first and second control gates insulated from the floating gates and located on opposite sides of the pillar.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 is a cross-section of floating gate transistors of an embodiment of the present invention;

[0013] Figure 2 is a plan view of a portion of a memory array of an embodiment of the present invention;

[0014] Figure 3 is a cross-section of the memory array of Figure 2;

[0015] Figures 4-15 illustrate a fabrication process of an embodiment of the present invention; and

[0016] Figure 16 is a block diagram of a memory of an embodiment of the present invention.

DETAILED DESCRIPTION OF THE DRAWINGS

[0017] In the following detailed description of the invention, reference is made to the accompanying drawings that form a part hereof, and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention.

[0018] To aid in the interpretation of the detailed description and claims that follow, the term "semiconductor substrate" is defined to mean any construction comprising semiconductive material, including, but not limited to, bulk semiconductive materials such as a semiconductor wafer (either alone or in assemblies comprising other materials thereon) and semiconductive material layers (either alone or in assemblies comprising other materials). The term "substrate" refers to any supporting structure, including, but not limited to, the semiconductor substrates described above. The term substrate is also used to refer to semiconductor structures during processing, and may include other layers that have been fabricated thereupon. Both wafer and substrate include doped and undoped semiconductors, epitaxial semiconductor layers supported by a base semiconductor or insulator, as well as other semiconductor structures well known to one skilled in the art.

[0019] In addition, as the structures formed by embodiments in accordance with the present invention are described herein, common semiconductor terminology such as N-type, P-type, N+ and P+ will be employed to describe the type of conductivity doping used for the various structures or regions being described. The specific levels of doping are not believed to be germane to embodiments of the present invention; thus, it will be understood that while specific dopant species and concentrations are not mentioned, an appropriate dopant species with an appropriate concentration to its purpose, is employed.

[0020] The term conductor is understood to also include semiconductors, and the term insulator is defined to include any material that is less electrically conductive than the materials referred to as conductors. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims, along with the full scope of equivalents to which such claims are entitled.

[0021] Finally, it will be understood that the number, relative size and spacing of the structures depicted in the accompanying figures are exemplary only, and thus were selected for ease of explanation and understanding. Therefore such representations are not indicative of the actual number or relative size and spacing of an operative embodiment in accordance with the present invention.

[0022] Non-volatile memory cells, as explained above, can be formed as floating gate transistors. While the area of the transistors can be reduced, the coupling ratio between the floating gate and the control gate is also reduced. Generally, as the gate-coupling ratio between the floating gate and the control gate decreases, the work voltage necessary to operate the memory transistor increases. As a consequence, the operational speed and efficiency of the flash memory decrease tremendously.

[0023] Some methods for increasing the gate-coupling ratio include: increasing the overlapped area between the floating and the control gate, reducing the thickness of the dielectric layer between the floating gate and the control gate, and increasing the

dielectric constant (k) of the dielectric layer between the floating gate and the control gate. Generally, to achieve an increase in the overlapped area between the floating and control gates and thus increase the gate-coupling ratio, the size of the floating gate has to be increased. However, this is not desirable for the demands of today's highly-integrated technologies.

[0024] Referring to Figure 1, a cross-section view of two floating gate transistors 100 according to one embodiment of the present invention is provided. The transistors includes a source 102, drain 104, floating gate 106 and a control gate 108. Unlike prior transistors, the source and drain regions of the transistors of the present invention are not located in a common horizontal plane of the memory device substrate. As explained below, the source and drain regions are fabricated in different planes such that a channel 110 of the transistor extends in a vertical direction. The floating gate and control gate also extend vertically into the substrate to control the transistor.

[0025] As shown in Figures 2 and 3, multiple floating gate memory cell transistors are fabricated in a substrate 200. Figure 2 is a plan view of a portion of a memory array and Figure 3 is a cross-section view of the array. The transistors share a common diffused source region 202, and the drain regions 204 are fabricated in a plane above the source region. Two adjacent transistors have independent floating gates 206 and 208 fabricated in trenches. A control gate 210 is fabricated between memory cells and is coupled to a word line 212 that runs parallel to the length of the control gates. Described in an alternate way, a pillar of silicon has two transistors fabricated therein. A top of the pillar has a common drain and two source regions are formed toward the bottom of the pillar. Floating gates are formed on opposite sides of the pillar. Each floating gate has a separate control gate. The sources and control gates extend in general parallel directions.

[0026] The features of one embodiment of the present invention can be described in greater detail with reference to a method of fabricating the transistors. The method is described in sufficient detail to provide an understanding of one method to form

transistors of the present invention. It will be understood by those skilled in the art that all process steps are not described in detail herein, and that extra steps or modifications of the steps may be needed depending upon the integrated circuit design and manufacturing equipment.

[0027] Figure 4 is a perspective cross-section view of a semiconductor substrate. The substrate has a layer of oxide (not shown) and a layer of Nitride 220 formed on the top surface of the substrate. The Nitride layer is then patterned and etched in the X direction to form shallow trenches 222 into the substrate. A layer of thin oxide is then grown and the trenches are filled with oxide 224. A chemical-mechanical planarization (CMP) operation is then performed to remove oxide above the Nitride layer. As such, the trenches are replaced with oxide isolation rows extending in the X direction; see the side view illustrated in Figure 5.

[0028] Referring to Figures 6 and 7, the top surface is then patterned and etched to form trenches running perpendicular to the oxide isolation rows. As a result, pillars 240 of substrate (active) material and oxide isolation 224 are formed. Floating gates and control gates will be formed in the trenches, as described below. As such, two transistors are formed using one active pillar. Each transistor has a separate control gate formed on opposite sides of the pillar.

[0029] A tunnel oxide layer 242 is grown over the pillars and a layer of polysilicon 244 is deposited over the oxide, see Figure 8. This layer of polysilicon is used to form the floating gates. Optionally, a layer of silicide can be combined with the polysilicon to form the floating gates. Referring to Figures 9 and 10, the polysilicon is patterned and etched to define the vertical edges of the floating gates 246. The horizontal edges of the floating gates are formed during the etch, or using a second subsequent etch operation. An implant operation is performed to form the source regions 250 in the bottom of the trenches 230. As such, the sources are located vertically beneath the control gate locations.

[0030] Referring to Figures 11-13, an oxide-nitride-oxide (ONO) thin film dielectric 252 is deposited to insulate the second floating gates and a polysilicon layer 254 is deposited. A layer of tungsten 256 is deposited in the groove 258 of the polysilicon. A CMP operation is then performed to remove portions of the tungsten and second polysilicon to form the control gate 260. The cross-section of the control gate, see Figure 12, resembles a U-shape polysilicon channel, or rail, with a center of tungsten.

[0031] In subsequent processes, the nitride material 220 is stripped and the drain regions 262 are formed through an implant operation at the top of the pillars. An insulating layer 270, such as boro-phospho-silicate glass (BPSG), is deposited and processed to form contact openings. The contact openings are filled with a conductor or semiconductor to form contacts 274. A metal layer is then formed and processed to create bit lines 276 coupled to the drain regions via the contacts 274, see Figure 14. Figure 15 illustrates the source regions 250 and source contacts 280 that provide electrical contact between the source and a source line 282.

[0032] Figure 16 is a functional block diagram of a memory device 400, of one embodiment of the present invention, which is coupled to a processor 410. The memory device 400 and the processor 410 may form part of an electronic system 420. The memory device 400 has been simplified to focus on features of the memory that are helpful in understanding the present invention. The memory device includes an array of memory cells 430. The memory cells are non-volatile floating-gate memory cells with vertical floating gates. The memory array 430 is arranged in banks of rows and columns.

[0033] An address buffer circuit 440 is provided to latch address signals provided on address input connections A0-Ax 442. Address signals are received and decoded by row decoder 444 and a column decoder 446 to access the memory array 430. It will be appreciated by those skilled in the art, with the benefit of the present description, that the number of address input connections depend upon the density and architecture of

the memory array. That is, the number of addresses increases with both increased memory cell counts and increased bank and block counts.

[0034] The memory device reads data in the array 430 by sensing voltage or current changes in the memory array columns using sense/latch circuitry 450. The sense/latch circuitry, in one embodiment, is coupled to read and latch a row of data from the memory array. Data input and output buffer circuitry 460 is included for bi-directional data communication over a plurality of data (DQ) connections 462 with the processor 410. Write circuitry 455 is provided to write data to the memory array.

[0035] Command control circuit 470 decodes signals provided on control connections 472 from the processor 410. These signals are used to control the operations on the memory array 430, including data read, data write, and erase operations. The flash memory device has been simplified to facilitate a basic understanding of the features of the memory. A more detailed understanding of internal circuitry and functions of flash memories are known to those skilled in the art.

CONCLUSION

[0036] A floating gate transistor has been described that includes source and drain regions that are fabricated on different horizontal planes. A floating gate and a control gate are fabricated vertically to control current conducted through the transistor. The control gate is coupled to a word line that is formed with the control gates and extends in a common horizontal direction.

[0037] Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement, which is calculated to achieve the same purpose, may be substituted for the specific embodiment shown. This application is intended to cover any adaptations or variations of the present invention. Therefore, it is manifestly intended that this invention be limited only by the claims and the equivalents thereof.